

A Content Based Assessment of the Relative Quality of Leading Accounting Journals

Abstract

This analysis advances faithful representation of statistical evidence as a substantive basis for assessing accounting journal research quality. The analysis builds upon recent work by Cready et al. (2019) indicating that accounting research articles commonly misrepresent null outcomes in their abstracts. Our analysis exploits this reporting deficiency to objectively assess journal reporting quality. The analysis determines misrepresentation rates for five leading general interest academic accounting journals based on direct review of article abstract contents. While all five of these journals commonly publish articles containing such misrepresentations, the relative frequencies with which they do so differ considerably. Moreover, the resulting rankings vary from those commonly reported in existent accounting journal quality and impact assessments. The analysis also finds that financial and archival studies are less prone to statistical evidence misrepresentation while audit and experimental studies are more prone to engaging in such misrepresentation.

1. Introduction

An extensive literature exists addressing accounting journal quality. Much of the work in this area takes the form of either: (1) surveys of knowledgeable individuals—generally accounting scholars or academic coalitions (e.g., collections of accounting faculties such as accounting departments); (2) measurement of journal usage (e.g., article downloads) or impact (e.g. article citations). Such analyses are argued to inform faculty performance evaluation and promotion decisions, guide faculty research targeting, and facilitate program and institution research quality assessments. (Benjamin and Brenner, 1974; Brown, 2003; Barrick, Mecham, Summers, and Wood, 2017). Of particular relevance for this analysis, Brown (2003) further argues that they facilitate journal self-improvement efforts. That is, they provide feedback to journals about the quality of their product which in turn guides them to quality improving actions.

Survey-, usage-, and impact-based approaches to assessing quality do not map rankings to specific identifiable characteristics of journals or articles. Hence, while they do provide useful overall insights regarding quality attainment, they do not directly measure attributes that drive research quality. Moreover, because they lack a direct connection with article and journal content, they are open to influences that move the measure apart from content quality. In the citation-based assessment literature, for example, there is widespread concern regarding spurious or even contrarian sources of citations such as negative references, self-citations, gratuitous citations, etc. In the survey-based literature, commonly raised concerns include qualifications of individuals responding to the surveys, personal experience with journals, and familiarity with the full set of journals being evaluated.

Uniquely, our analysis proposes and implements a salient content-based measure of journal quality. The measure flows from recent work by Cready et al. (2019) on the widespread non-compliance of academic accounting articles with principles for the conduct, interpretation, and reporting of tests of statistical significance recently promulgated by The American Statistical Association (ASA). Cready et al. documents that articles routinely misrepresent statistically insignificant test outcomes (“null outcomes”) as a *per se* basis for concluding that tested null effect values (typically zero) are true. Such interpretations violate multiple principles for the conduct of statistical analysis promulgated in the *ASA Statement on P-Values and Statistical Significance* (Wasserstein and Lazar, 2016).¹ Consequently, from a financial accounting perspective of what constitutes high quality reporting, they do not “faithfully represent” such outcomes.

Our measure is of particular relevance as a quality assessment alternative to perception-based measures given the concerns that led to the promulgation of the *ASA Statement*.² Specifically, a widespread tendency of empirical researchers to over-interpret statistical test of hypothesis evidence. These concerns are, for instance, clearly seen in the Wasserstein, Schrim, and Lazar (2019) restatement of them in the form of a set of “don’ts”. Each of these “don’ts” directly addresses specific ways that researchers misrepresent p-value evidence by overstating its certainty and effect size implications. Hence, one way of viewing our analysis is as a direct assessment counterbalance to research impact enhancements obtained from unjustified over-interpretations of statistical evidence.

¹ See Amrhein, Greenland, and McShane (2019), Wasserstein, Schrim, and Lazar (2019), and Cready (2019) for recent definitive expositions addressing the inappropriateness of conclusive null outcome interpretations.

² Traditionally surveys are commonly associated with reflecting perceptions while usage-based approaches are not (Beets et al., 2015). However, at a broader level the choice to use or cite a journal article is also influenced by perceptions and, as is particularly relevant for our analysis, perceptions possibly determined by article misrepresentations of the certainty or effect size importance of p-value evidence.

Our study's general approach of using relevant authoritative reporting principles as a basis for assessing quality means that the analysis has much in common with the use of financial report restatement levels as a measure of firm reporting quality. Financial report restatements reflect objectively determined material reporting errors. Firms with higher restatement levels are typically taken as exhibiting lower reporting quality while firms with lower restatement levels are taken as exhibiting higher reporting quality. Similarly, we take journals with objectively determined higher null outcome mis-representation rates as providing lower quality research reporting.

Our assessment considers publications from the 2013 to 2017 time period from five journals: *Contemporary Accounting Research (CAR)*, *Journal of Accounting and Economics (JAE)*, *Journal of Accounting Research (JAR)*, *Review of Accounting Studies (RAST)*, and *The Accounting Review (TAR)*. Null outcome misrepresentation levels are disturbingly high in all five of these journals. When we focus on articles with abstracts containing misleading “precisely conclusive” interpretations (Cready et al., 2019), misrepresentation levels range from 11.3% to 24.4% of p-value reporting articles. When we also include less severe forms of misrepresentation in our measures, these levels rise to a range of 19.4% to 32.8%. *JAE* has the lowest misrepresentation rates, which is in line with its typical position as either the highest or second-highest impact factor journal in accounting. *JAR*, which is typically *JAE*'s closest rival in almost every accounting journal quality ranking exercise, has the highest misrepresentation rates. Hence, our evidence is consistent with those ranking approaches that place *JAE* as a top accounting journal. It also suggests that the case for *JAR* as a leading accounting journal must rest on aspects of its content other than the degree to which it provides forthright representations of examined evidence in the abstracts of its articles.

We also evaluate null outcome reporting in four accounting research sub-disciplines: Audit, Financial, Managerial, and Tax. Audit articles are the most prone to engaging in null outcome misreporting. Audit article abstracts contain inappropriate precisely conclusive null outcome interpretations 24.4% of the time. In contrast, financial article abstracts contain such language only 14.5% of the time. Similarly, when we partition articles based on whether they employ archival or experimental methods we find that experimental study abstracts misrepresent null outcomes using precisely conclusive language 22.5% of the time as compared to a 16.3% misrepresentation rate for archival analyses.

2. Journal Quality Assessment Perspectives

2.1 Indirect Assessment Perspectives

The majority of existent research assessing the quality of accounting academic journals relies on surveys of research knowledgeable user groups such as the general population of accounting academics, accounting department heads, and targeted accounting academics sub-populations such as academics working in specific geographical areas (e.g., Australia, Canada, UK), at a certain type of research differentiated institution, or of a certain rank (e.g., Howard and Nikoli, 1983; Hull and Wright, 1990, Benjamin and Brenner, 1974). After excluding finance journals, TAR, JAR, and *Accounting, Organizations, and Society* (AOS) typically occupy the top spots in the earliest such studies. In the late 1980s *JAE*, first published in 1979, also emerges as a front-runner in these studies. In a recent compilation of mostly survey-based rankings published since 2000, Bujaki and Mcconomy (2017) present an integrated “top 10” ranking of : (1) *JAR*; (2) *JAE*; (3) *TAR*; (4) *CAR*; (5) *Auditing: A Journal of Practice and Theory*; (6) *AOS*; (7)

Journal of Accounting, Auditing, and Finance; (8) *Journal of the American Taxation Association*; (9) *RAST*; and, (10) *Journal of Management Accounting Research*.

Citation-based analyses are also employed to assess journal quality. This approach is advanced by Dyckman and Zeff (1984) and Brown and Gardner (1985) as a more objective means than surveys for assessing the “impact” of accounting journals on “contemporary accounting research.”³ In the current publishing environment their most prominent role is as top line need-to-know content provided on the opening page of journal websites. In recent years they appear in academic journal quality assessments largely in the form of variations on conventional assessment citation metrics (e.g., the two-year journal impact factor readily available from *Journal Citation Reports*). For instance, Chan et al. (2009) evaluate weighting citation counts in accounting research sub-fields by relative numbers of dissertation citations and Barrick, Mecham, Summers, and Wood (2017) evaluate the degree to which an article is cited by articles from research areas outside of its own area.

2.2 Direct vs. Indirect Assessment

Survey and citation approaches are both indirect measures of journal quality. They do not measure primitive quality drivers derived from the actual content of journal articles based on examinations of the methods and procedures by which articles are selected and edited or due to the general management practices followed by journals. Instead, they address plausible consequences of journal quality based largely on perceptions of knowledgeable individuals. High quality journals should be more highly regarded by scholars in the field and publish articles that are read more often and cited more frequently by such scholars. However, a journal may perform better than another based on these measures simply because it is better known or has a captive

³ McRae (1974) is the first study to our knowledge that employs citation analysis in an accounting research context. His analysis, however, does not employ citations as a basis for ranking or assessing journal impact or quality.

audience (e.g., major association journals). Or, because it concentrates on publications in widely followed research areas.⁴ Journals may even enhance their standing by focusing on moving the measures apart from changing the fundamental journal content by means of questionable citation practices, marketing campaigns, and, as is relevant to this analysis, proclivity for providing perception-enhancing interpretations and conclusions in articles that lack substantive foundation in provided analyses.

In contrast to the outcome focus of surveys and citations, direct assessment focuses on observed journal content and operating procedures. That is, core attributes that high quality research should or should not exhibit are identified and measured directly, not indirectly. Hence, the approach is generally robust to many of the concerns associated with indirect perception-based measures. Direct content measures can be produced for any journal entirely apart from whether it is an association journal, is covered by a citation service, or is selected for inclusion in a given survey. They are also not subject to spurious influences arising from factors such as audience size, legacy, familiarity, historical standing, and unfounded perception-enhancing description of article content.

Direct assessment measures also directly identify what is being done right and what is being done wrong. For instance, *ceteris paribus*, a journal publishing a high percentage of articles containing plagiarized material is likely producing lower quality research than a journal containing only articles that are plagiarism-free. Similarly, given agreement on the general principle that a peer reviewed version of an article is of higher quality than a non-peer-reviewed version, it readily follows that a journal publishing peer reviewed material is exhibiting higher academic research quality than a journal that is not employing peer review. Moreover, unlike indirect

⁴ Multiple studies acknowledge and to varying degrees attempt to rectify, ranking biases attributable to the dominant role of financial topic area articles in accounting research. (e.g., Bonner et al., 2006).

assessment, the metric itself also directly speaks to what needs to be done to improve quality. For instance, if plagiarism is present, reducing it will improve journal quality. If peer review is absent, implementing it will improve quality.⁵

While direct assessment has much to recommend itself relative to indirect assessment, it also has key limitations. First, any direct assessment metric requires identifying a compelling rationale for taking it as a valid measure of quality. The difficulty of such identification cannot be overstated. Academics, at best, tend to agree to disagree about what constitutes quality research efforts (Brinn, Jones, and Pendlebury, 1996). That said, however, there are some attributes about which there is widespread agreement. For instance, plagiarism (given agreement on what constitutes plagiarism) is widely thought of as, at best, signifying poor quality. Similarly, *Clarivate* provides several journal “publishing standards” it applies in a journal’s application for listing in its various citation indices (e.g., the *Social Science Citation Index*). These include peer review, presence of acknowledgements (reflecting the article being read and commented on by non-authors prior to publication), and ethical publishing practices. All of these are standards that most scholars seem likely to agree upon as factors relevant to the production of quality research.⁶

A second limitation of direct measures is that they are narrowly focused. That is, they measure a specific attribute that quality research should either exhibit or not exhibit. As research

⁵ Brinn, Jones, and Pendlebury (1996) survey accounting academics regarding the acceptability of ten questionable research practices. Three practices were widely viewed as unacceptable: (1) replication another researcher’s work without attribution; (2) failing to appropriately recognize colleague contributions to a study; (3) simultaneous submission of a paper to multiple outlets. Two others were identified as “marginally unacceptable”: (1) inclusion of a minimal contributor as a co-author; (2) publishing highly overlapping studies at two separate journals.

⁶ Bean and Bernardi (2005) and Matherly and Shortridge (2009) examine the journal quality determinant issue empirically. They identify journal age, journal search engine visibility, article page length, journal audience (academic or practitioner), *Social Science Citation Index* covered, and submission fee as predictors of perceived journal quality. These relations, however, reflect on average relations that are, in some cases, heavily perception driven. That is, they do not reflect primitive drivers of article quality.

quality is determined by an array of attributes, many of which are inherently subjective, a single metric that happens to be objectively measurable should not be taken as reflecting the overall research quality of a journal. It is simply a measure of quality on a specific dimension. Hence, for example, just because a journal avoids publishing articles containing plagiarized material does not imply that it is a high research quality journal. Rather, it has simply met a necessary condition for being possibly taken as being a high-quality journal. Similarly, the assessment relevance of any such metric also depends on the fundamental importance of what it measures. Making up evidence or plagiarism are of obvious importance for quality. On the other hand, the substantive research quality relevance of article compliance with things such as generally recommended formatting guidance or avoiding gratuitous citations are less obvious.⁷

2.3 Null Outcome Misrepresentation as a Direct Assessment Metric

The direct assessment approach we advance concerns the extent to which journals publish articles with materially misleading descriptions of statistical evidence. The assessment merits of this criterion stem directly from the financial reporting notion of “faithful representation.” Faithful representation is identified in *Statement of Financial Accounting Concepts Number 8* as one of three “Fundamental Qualitative Characteristics” of financial information. In the Statement’s language, a “perfectly faithful representation” of a phenomenon has three characteristics. “It would be *complete*, *neutral*, and *free from error*.” In our opinion, faithful representation is a broad reporting principle that the vast majority of scholars view as a reasonable expectation that quality research efforts should aspire to exhibit.

⁷ From an indirect measurement perspective, however, one might reasonably argue that the degree to which journals or articles follow good practices in small things likely signals the degree to which they follow good practices in more important things.

The specific statistical misreporting construct we address concerns misrepresentation of “null outcomes” in article abstracts. Cready et al. (2019) define null outcomes as “instance(s) where a non-directional test of a (null) hypothesis results in p -values that are larger than what is required to sustain a rejection inference.” Fundamentally, a null outcome in and of itself implies that the examined evidence is compatible with both null and salient alternative effect size hypotheses. (See, in particular, discussion accompanying Principle 6 of *The ASA Statement on P Values and Statistical Significance*.) That is, from a hypothesis testing framework, the examined empirical evidence does little to resolve whatever prior uncertainty exists regarding a tested hypothesis or effect value of interest. A faithful representation of a null outcome should convey uncertainty, not certainty, with respect to the implications of the examined empirical evidence for the relevant research question.

The null outcome interpretation evidence reported in Cready et al. (2019), however, indicates that accounting publication abstracts commonly present null outcomes as compelling evidence of the truth of the associated null hypotheses, not as evidencing uncertainty as to whether the evidence better supports this null or its companion alternative hypothesis(es).⁸ That is, articles routinely provide distinctly unfaithful representations of null outcome evidence. Moreover, contextually these misrepresentations commonly substantively enhance nominal “contributions” of articles. Statistical evidence implying nothing more than uncertainty about the presence or absence of an effect morph into “surprising” or “important” “findings” that no effect whatsoever is present. Essentially, null outcome misrepresentation serves as a clever rhetorical

⁸ Other analyses document widespread misrepresentation of null outcomes in sociology (Bernardi, Chakhaia, and Leopold; 2017), biology (Fidler, Burgman, Cumming, Buttrose, and Thomason; 2006), psychonomics (Hoekstra, Finch, Kiers, and Johnson; 2006), and neuropsychology (Schatz, Jay, McComb, and McLaughlin; 2005).

device for advancing unfounded notions as to why article contents are interesting, important, and impactful.

At a nominal level, the case for taking null outcome misrepresentation as a form of material descriptive deception that respectable research efforts should avoid is unambiguous.⁹ Cready (2019) identifies it as “statistical malpractice.” In their assessment of *p*-value misuse in the *Psychometric* literature Hoekstra, Finch, Kiers, and Johnson (2006) null outcome misrepresentation is the only offense that rises to the level of being a “serious mistake.” Amrhein et al. (2019) in a *Nature* editorial assert that such misrepresentations are “ludicrous” and “nonsensical.” Nevertheless, despite its face value inappropriateness, there is a line of thought that such misrepresentation, while superficially wrong, is truly inconsequential as a matter of practice. This line of thinking views such descriptions as harmless exaggeration or simplification that facilitates clear concise presentations of empirical evidence (Cready, 2019). It presumes that journal readers see through the “certainty that no effect is present” language in articles and correctly read it as truly conveying broad uncertainty about effect size and direction. That is, null outcome misrepresentation falls into some sort of “wrong but harmless” category of statistical malpractice.

While superficially attractive, there are several compelling reasons for questioning this no harm perspective. First, it is at odds with the ASA statement’s self-identified purpose of providing “a few select principles that could improve the conduct or interpretation of quantitative science, according to widespread consensus in the statistical community.” The principles in the ASA statement broadly address what they perceive as harmful violations of elementary sound

⁹ Notably, the role of the use of null outcomes as a basis for asserting contingent liability immateriality was recently litigated. In *Matrixx Initiatives, Inc. v. Siracusano* (2011) the United States Supreme Court unanimously “rejected the argument that the risk of side effects from a pharmaceutical product can never be “material” under the securities laws so long as the risk is not known to be statistically significant.” (Ericson and Brandt, 2011)

practices, violations commonly encountered in statistical applications in the broader research community. There is no basis that we are aware of for taking null outcome misrepresentation as some sort of special exception to the statistical malpractice concerns that led to the promulgation of the ASA statement. Indeed, the fact that the Amrhein et al. (2019) discussion focuses extensively on null outcome misinterpretation in advancing its case against the use of “statistical significance” strongly indicates that it is exactly the sort of sound practice violation that led to the promulgation of the statement.

Second, this no harm perspective is inherently self-contradictory. If null outcome misrepresentation does not impact reader beliefs, understandings, and assessments, why do authors seeking to publish their work routinely engage in something that can only diminish reader perceptions of the quality of their work? Moreover, why do articles almost never faithfully represent null outcome evidence (Cready et al., 2019)? Such behaviors make a good deal of sense if misrepresentation works. Articles representing resolutions of issues based on evidence generally fare far better in review processes than articles that portray uncertainty about definitive implications of empirical evidence. Alternatively, if misrepresentation fools no one, these patterns imply that authors have an inexplicable aversion to faithful representation of null outcome evidence that is facilitated by fully knowledgeable reviewers and editors with a remarkable tolerance for careless interpretations of such evidence.¹⁰

Finally, if there is some sort of general awareness of the fact that null outcomes are fundamentally identifying an uncertainty of the implications of evidence for relevant hypotheses,

¹⁰ The no harm perspective presupposes that all relevant players truly **know** what is going on (which is not at all the same as being vaguely aware of what is going on) and so are not misled by nominally deceptive representations. Knowingly misrepresenting evidence, absent a certain belief that others are not typically misled by such misrepresentation, is ethically problematic. At a substantive level, it differs from outright fabrication of evidence only in that there is a possibility that knowledgeable readers may or should not be misled based on examining the totality of the presented material.

why is there an almost complete absence of substantive engagement with the underlying precision of the evidence underlying null outcomes by articles reporting such outcomes? As a practical matter, engagement with such precision is the only path to drawing useful insights from null outcome evidence. Yet, it is rarely found for null outcomes reported by accounting research articles (Cready et al. 2019). Indeed, in many instances articles report null outcomes where, as evidenced by relevant confidence intervals, the underlying evidence is far more compatible with highly material contrarian alternative hypotheses than it is with asserted “no effect at all” interpretations? (e.g., table 5 of Cready et al. 2019).

Collectively, we take the preceding arguments as casting considerable doubt on the conjecture that the accounting discipline practices null outcome misrepresentation from a position of full practical knowledge and understanding. Absent such full knowledge, it necessarily follows that null outcomes misrepresentations commonly deceive. Hence, we think there is considerable merit to taking the degree to which a journal publishes such deceptions as a substantive measure of its quality.

2.4 Null Outcome Misrepresentation Measurement

Our analysis evaluates null outcome misrepresentation in five widely recognized academic accounting journals (*CAR*, *JAE*, *JAR*, *RAST*, and *TAR*) over the 2013 to 2017 time period. As the analysis requires textual examination of individual article abstracts, we, as a matter of practicality, can only assess a limited number of journals. Given these inherent constraints we opt to focus on a set of highly regarded journals that are as comparable as possible. Hence, we exclude specific area-oriented journals such as *Auditing* and non-North American journals. Given these exclusions, the five selected journals are, in our opinion, obvious candidates. *CAR*, *JAE*, *JAR*, and *TAR* appear in the top 5 of almost every accounting journal

ranking exercise (exclusive of any included finance journals) produced since 1990.¹¹ RAST does not appear in that many rankings due to its relatively recent inception (1996). However, in their compilation of journal rankings, Bujaki and Mcconomy (2017) place RAST as number 6 in their list of “discipline journals”, just behind AOS.

Abstracts for every article published (1,168 in total) by each of these journals were read.¹² Based on these readings, with further text examination conducted as necessary, every article was classified with respect to research topic area (Audit, Financial, Managerial, Tax, and Other) and research method (Archival, Experiment, Other)¹³. Our analysis focuses on the subset of these articles (1,030) that employ statistical significance based empirical inference. We also identify candidate of abstract level null outcome interpretations based on these readings. An interpretation was identified as pertaining to a null outcome only if we were able to confirm that the abstract language corresponds to high p-value evidence presented or discussed in the reporting article’s text. This initial abstract review process was conducted independently by two of the study authors. Classification disagreements were settled by consultation by all three authors.

After completion of the initial text analysis, we further classified the identified null outcomes based on how article abstracts interpret them. We followed the five null outcome

¹¹Bonner, Hesford, Van der Stede, and Young (2006) provide a comprehensive summary and analysis of journal ranking studies through 2004 while Bujaki and Mcconomy (2017) summarize the post-2000 literature.

¹² We do, however, rely on the Cready et al. analysis with respect to articles published in *TAR* in 2017 and 2017.

¹³ For method classifications articles were classified as archival or experimental only if they also employed statistical significance-based assessment. 140 of the 1,168 articles, mostly analytical studies, examined do not employ significance-based empirical assessment in meaningful fashion. Also, survey-based empirical analyses are classified as archival and information system topic area studies are classified as managerial. Articles deemed as addressing two distinct research areas are classified by the following hierarchy: tax if one of two areas is tax; audit if one of the two areas is audit and the other areas is not tax; managerial. Some abstracts also contain multiple null outcome interpretations falling into different categories. In these instances an article was classified PC if it contained a PC interpretation, it was classified as SC if it contained an SC interpretation and no PC interpretations, and GC if it contained a GC interpretation and no PC or SC interpretations.

categories from Cready et al. (2019) in doing this: (1) precisely conclusive (PC); (2) generally conclusive (GC); (3) selectively conclusive (SC); (4) arguably conclusive (AC); and, (5) non-conclusive (NC).¹⁴ Articles not reporting null outcomes in their abstracts were classified as non-null (NN). Again, two authors independently coded each of the previously identified null outcomes into one of these categories. In the few instances where initial codings differed, the two coders each further examined the relevant article discussions and came to a final agreement on a final classification. The final category counts were: 175 articles with PC interpretations; 62 with GC interpretations; 28 with SC interpretations; and, 1 with an NC interpretation. The NC article, Joo and Chamberlain (2017), was published by *CAR*. Its publication is, in our opinion, a notable achievement. It stands alone as the one instance where an article and journal can unambiguously claim to have faithfully represented null outcome evidence in an article abstract.¹⁵

We construct three summary measures of null outcome misrepresentation based on the preceding abstract classification exercise. The first measure is the percentage of a relevant article collection K that provide questionable null outcome representations in their abstracts. It is calculated as:

$$PQ_K = \frac{\text{Number of articles in collection K containing PC, GC, or SC null outcome interpretations}}{\text{Total Articles in Collection K}} .$$

Article collections for which we calculate this metric include: (1) all p-value reporting articles in the five journals we study; (2) all p-value reporting articles published by a journal; (3) all p-value

¹⁴ In those instances where abstracts address multiple null outcomes, we coded the article as providing a PC abstract if any of its interpretations met the PC criteria and, for those articles without PC interpretations, GC when abstracts contain both GC and SC interpretation.

¹⁵ Cready et al. (2019) identify two articles that provide NC interpretations of null outcomes. However, these interpretations are found in the article text discussions, not in their abstracts.

reporting articles addressing a specific research area; and, (4) all p-value reporting articles using a specific research method.

GC and SC interpretations, as discussed in Cready et al., are possibly reasonable conditional on what other information and analysis an article brings to bear to interpret them. In particular, GC interpretations are arguably appropriate when a descriptive-oriented case is made for the effect size being inconsequential or small. SC interpretations are defensible if an article provides contextual discussion clearly recognizing that while the evidence is compatible with the hypothesized null effect size(s), it is also compatible with effects sizes that are not at all consistent with the postulated null hypothesis.¹⁶

Our second measure addresses the arguable aspects of including GS and SC interpretation as lacking representational faithfulness by focusing exclusively on PC classified articles. PC articles are never faithful representations of null outcomes. By definition, they advance the idea that a conjecture has been somehow proven by the examined evidence when, in fact, the underlying evidence is inherently inconclusive regarding the truth or falsehood of the conjecture in question. Our second measure, PPC, focuses on PC representations of null outcomes. It is calculated as:

$$PPC_K = \frac{\text{Number of articles in collection K containing PC null outcome interpretations}}{\text{Total articles in collection K}} .$$

One drawback to the PQ and PPC measures is that their magnitudes also depend on a journal's tendency to publish articles presenting null outcome evidence. Given that null

¹⁶ Cready et al. (2019), however, find little support for articles doing either of these things in their text discussions. Similarly, in our own somewhat less intensive review of the set of GC and SC identified articles found little evidence of substantive descriptive assessments of effect size magnitudes by GC articles or of balancing interpretative discussions in SC articles.

outcomes are typically misrepresented when an article chooses to report them, it follows that journals favoring null outcome reporting may perform poorly with respect to these two metrics simply because they publish more null outcome reporting articles. We address this concern with our third measure, PNPC, the percentage of a journal’s null outcome reporting articles that provide PC interpretations. This measure is calculated as:

$$\text{PNPC}_K = \frac{\text{Number of PC null outcome reporting articles in collection } K}{\text{Total number of null outcome reporting articles in collection } K}$$

PNPC directly reflects the degree to which articles that do report null outcomes elect to provide the most extreme form of unfaithful representation interpretations for them. Ideally, we would complement this metric with a second one where the numerator is instead based on the presence of any sort of questionable null outcome representation (i.e., it would include GC and SC interpretations). However, given the actual evidence, such a metric would be pointless. The numerator would almost always equal the denominator since there is only one instance where an article avoids providing a possibly tainted interpretation for an abstract-reported null outcome.

3. Empirical Evidence

3.1. Preliminary Analyses

An important dimension of our analysis is the degree to which our representational faithfulness assessments differ from or confirm conventional survey and citation-based outcomes. Accordingly, Table 1 reports relevant annual citation-based “impact” assessments of journal quality for the 2013 to 2017 time period for the five journals examined in our analysis. The first metric, reported in panel A, is the conventional citation impact factor provided by *Journal Citation Reports*. It consists of the number of citations in the indicated year attributable to articles published in the prior two years. Hence, the 2018 column reflects citations to articles

published in 2017 and 2016 (which are the last two years that we evaluate abstract reporting of null outcomes). The second metric is the “SCImago Journal Rank” follows the same form as the *JCR* impact factor, except that it weights citations by the importance of the journal that is providing the cite. These rankings are publicly available at <https://www.scimagojr.com/>. Both metrics paint a similar picture of the relative journal standings over the time period we study. JAE and JAR typically take the top two spots with JAR holding the top spot in 2018 (cites to publications in 2017 and 2016) and 2017 (cites to publications in 2016 and 2015) and JAE holding the top spot in 2016 and 2015. TAR consistently holds down the third spot while CAR generally comes out ahead of RAST. The ranking picture portrayed here differs from post-2000 journal quality rankings, as summarized in Bujaki and Mcconomy (2017), only in that survey based assessments commonly place *TAR* in the number 1 position and almost always place *JAE* below *JAR*.

Tables 2 and 3 summarize relevant cross-tabulations of articles produced by the initial abstract reviews. Panel A of table 2 presents the numbers of articles published in major sub-accounting sub-fields by journal. Not surprisingly, the financial field dominates, with 62.2% of the articles published across the five journals falling into this category. There is, however, considerable variation in financial topic proclivity across journals with RAST achieving a high value of 79.1% and TAR, consistent with its role as an association journal, coming in with a low value of 51.9%. JAE, JAR, and RAST all seem relatively disinclined to publishing audit articles as fewer than 10% of their articles fall into this category. In contrast, over 20% of the articles in CAR and TAR address auditing. There is a similar, but less pronounced, divide in the managerial arena, with CAR and TAR again being more inclined to publishing managerial articles relative to

JAE, JAR, and, in particular, RAST. Finally, all of these patterns persist when the sample is restricted to the 1,030 articles employing statistical significance testing methods (panel B).

Table 3 reports how archival and experimental significance testing designs divide by research area (panel A) and by journal (panel B). The panel A distributions indicate that experimental designs are rarely employed by financial (5.8% of the time) and tax (1.2% of the time) analyses. In contrast, managerial analyses employ experimental methods 30.5% of the time, and audit analyses employ them 20.7% of the time. Preference for archival relative to experimental methods also varies considerably by journal. JAE and RAST almost never publish experimental studies. In contrast, the other three journals publish them at similar rates—14.7% for CAR; 14.0% for JAR; and 16.9% for TAR.

3.2. Journal Assessments

Table 4 presents abstract null outcome misrepresentation rates for the five journals. On average, 25.7% of p-value reporting articles in these journals misrepresent null outcomes in their abstracts. 17% of p-value reporting articles, in fact, present null outcomes in highly misleading terms. They provide PC interpretations of null outcomes proclaiming that the evidence clearly demonstrates the absence of an effect when, in fact, the underlying evidence is quite compatible with its presence. These misrepresentation rates are also substantial across all five journals. All of the rates exceed 10%. At the individual journal level, JAE and JAR, the two journals that typically come out at the top of citation-based journal quality assessments such as those we report in table 1, differ markedly. JAE has the lowest percentage of articles that misrepresent null outcomes in abstracts—19.4%. JAR has the highest such percentage—32.8%, over 50% higher than the *JAE* rate. When we focus on PC misrepresentations, the rates are 24.4%

versus 11.3% with the *JAR* rate again being the highest of the five journals we consider.

Collectively, the evidence favors *JAE* as the (relative) quality leader in this set of journals. They also suggest that the case for viewing *JAR* as a research quality leader must rest on dimensions of its content other than the degree to which the articles it publishes faithfully represent null outcome evidence.

The other three journals (*CAR*, *RAST*, and *TAR*) substantially outperform *JAR* in terms of avoiding null outcome misrepresentations in their abstracts, but they also fall well short of the *JAE* standard. Unlike our table 1 citation-based quality rankings, there is little support here for taking *TAR* as possessing the highest quality among these three journals. *RAST*, in fact, outperforms it on all three measures and *CAR* outperforms it on one of the three. More critically, the performance differences are simply not very large across these three journals, broadly consistent with taking them as largely indistinguishable from one another in their handling of null outcome reporting.

One weakness of the table 4 analysis is that it does not directly control for differences in content mix across journals. As is evident from the table 2 and 3 cross-tabulations, substantial differences exist across these five journals in terms of the mix of research area and methods they typically publish. Null outcome incidence rates or attention to representation faithfulness likely differs by discipline or method. Hence, what we attribute to journal level practices based on the table 4 values may actually reflect differences in the mix of content published by a journal. That is, for example, *JAE*'s strong relative performance may have more to do with its aversion to publishing articles employing experimental methods and its strong preference for articles addressing financial topics. We revisit the robustness of the table patterns to article mix in a robustness analysis presented in a subsequent section. In general, however, this analysis does not

contradict these implications for cross-journal differences in null outcome reporting quality.

3.3. Research Area and Method Quality Assessments

As our three measures of null outcome misrepresentation directly measure the level of misrepresentation present in any studied collection of articles, they are readily amenable to meaningful cross-area and cross-method assessment exercises. Table 5 reports two such exercises. Panel A reports our three misrepresentation metrics by the previously identified four major research areas: (1) Financial; (2) Managerial; (3) Audit; and (4) Tax. The evidence broadly suggests that financial articles do a somewhat better job of avoiding misleading representations of null outcomes in abstracts, reporting rates of 23.2% for the overall metric, 14.5% for the PC focused metric, and 62.5% for the relative metric (an inverse measure of misrepresentation). Audit article abstracts, on the other, contain null outcome misrepresentations 35.4% (broadly defined) of the time and 24.4% of the time they contain PC misrepresentations. Both of these values are considerably higher than the misrepresentation levels exhibited by the other research areas. This evidence is loosely consistent with the notion that journals that favor publishing financial reporting articles and disfavor publishing audit articles such as JAR, JAE, and RAST should, *ceteris paribus*, exhibit lower misrepresentation rates.

Panel B divides the articles by whether they employ archival or experimental methods. When misrepresentation is broadly defined, these misrepresentation rates are similar—25.9% for archival and 26.1% for experimental. However, they differ when we focus on the specific tendency to engage in PC misrepresentation. Experimental studies are more prone than archival studies to providing PC misrepresentations—22.5% vs. 16.3%. This tendency in favor of PC misrepresentation is particularly evident in the relative metric scores of 86.2% versus 62.8%.

When an experimental analysis opts to discuss a null outcome in its abstract, it almost always does so with a precisely conclusive assertion that no effect is present.

4. Robustness and Supplemental Analyses

We evaluate the potential confounding influences of variation in the mix of articles published by journals on our null outcome misrepresentation metrics by restricting the analysis to the subset of articles addressing financial issues using archival methods. 618 articles met this criterion. Table 6 reports values for our three null outcome misrepresentation metrics by journal for this sub-group. Given the table 5 evidence, this particular subset of articles should, in general, report and hence misrepresent null outcomes in abstracts less frequently than articles that do not fall into it. Consistent with this conjecture, overall null outcome misrepresentation percentages are lower than those reported in table 4. This difference is most pronounced for the use of PC interpretation for when reporting null outcomes in abstracts. The table 6 percentage is 57.3% as compared with 65.3% in table 4. The most notable journal level difference in table 6 relative to table 4 is the emergence of TAR as a plausible contender to JAE for the top spot in the rankings. TAR puts up misrepresentation percentages of 20.3% and 12.0% for the two overall frequency metrics. These are only slightly higher than the 17.7% and 10.6% JAE numbers. TAR actually slightly outperforms JAE (59.4% versus 60.0%) in terms of articles avowing the use of PC interpretations when discussing null outcomes in abstracts.

A final natural relevant question concerns what is happening over time with respect to the level of non-faithful representation of null outcomes in the accounting literature. As the table 4 evidence demonstrates overall misrepresentation levels are unacceptably high. But, are they improving? Or, are things getting worse? We provide some preliminary evidence on what might

be going on over time in table 7. This table reports misrepresentation rates by year for the five years covered in our analysis. The reported rates vary some from year to year, but there is no compelling evidence of any sort of trend to them. Of course, five years is a short time-series. Hence, a longer trend might prove more insightful.

5. Conclusion

Conventionally, journal quality assessment relies on indirect measures such as opinion surveys, download activity, and citation levels that address likely or expected research quality outcomes. This analysis advances measures of the degree to which a journal's articles faithfully represent statistical evidence as a complementary direct measure of research reporting quality. The direct nature of this measurement approach yields quality insights incremental to those gleaned from existent measures. Specifically, within a set of widely recognized leading accounting journals, substantive differences exist in null outcome misrepresentation propensities. Moreover, when these journals are ranked based on their tendency to publish such misrepresentations, the ordering is substantively different from that produced by more conventional perception-based rankings. Indeed, *JAR*, a journal that typically comes out on top in accounting journal rankings comes in last (by a considerable margin) based on the extent to which a journal's articles misrepresent null outcome evidence.

The analysis also finds that research efforts in the auditing field and those employing experimental methods are particularly prone to misrepresenting null outcome evidence. A likely explanation for these tendencies is that these are both low power research settings. Hence, the misrepresentation is a consequence of being more prone to encountering null outcomes. Importantly, this perspective simply provides an understanding for such higher misrepresentation

rates. It does not, at an absolute level, undercut the fact that misrepresentation is higher in these areas than in other areas. Finally, it is also noteworthy that experimental studies, relative to archival ones, tend to favor the most extreme form of misrepresentation (i.e., that the null is unquestionably true) over less extreme forms (i.e., SC and GC interpretations). This emphasis possibly indicates that such studies tend to compensate for weak power (due to small sample sizes) with aggressive interpretation of empirical evidence.

In understanding the various rankings presented here, it is important to reiterate that they only reflect a single quality dimension that happens to be directly measurable. Hence, they are appropriately interpreted as simply adding a degree of balance to a current scoring system that is largely perception driven. While avoiding null outcome misrepresentation or, more generally, overstatement of the certainty associated with statistical test outcomes certainly improves research quality, such aversion does not of itself make a journal or a research discipline high quality. Alternatively, however, a performance failure on this dimension is informative as to absence of quality. Moreover, assessing this absence is particularly useful when, as seems likely to be true here, the performance failure's effect is unfounded enhancement of perceived article quality.

References

- Amrhein, A., S. Greenland, and B. McShane. 2019 "Scientist rise up against statistical Significance," *Nature* 567, 305-307.
- Barrick, J.A., N.W. Mecham, D.L. Summers, and D.A. Wood, 2017, "Ranking accounting journals by topical area and methodology," working paper, November.
- Bean, D. and R. Bernardi, 2005. "Estimating the ratings of journals omitted in prior quality ratings." *Advances in Accounting Education* 7: 109-127.
- Beets, S. D., A.S. Kelton, and B.R. Lewis, 2015, "An assessment of accounting journal quality based on departmental lists," *Scientometrics* 102: 315-332.
- Benjamin, J., & V. Brenner (1974). "Perceptions of journal quality." *The Accounting Review* 49(2): 360-362.
- Bernardi, F., L. Chakhaia, and L. Leopold. 2017. "'Sing me a song of social significance': The (mis) use of statistical significance testing in European sociological research." *European Sociological Review* 33(1): 1-15.
- Bonner, S.E., J.W. Hesford, W.A. Van der Stede, and S.M. Young, 2006, "The most influential journals in academic accounting," *Accounting, Organizations, and Society* 31: 663-685.
- Brinn, T., M. J. Jones, and M. Pendlebury, 1996, "UK accountants' perceptions of research Quality," *Accounting and Business Research* 26(2): 265-278.
- Brown, L., 2003, "Ranking journals using social science research network downloads," *Review of Quantitative Finance and Accounting* 20: 291-307.
- Brown, L., and R.J. Huefner, 1994, "The familiarity with and perceived value of accounting journals: Views of senior accounting faculty in leading U.S. MBA programs," *Contemporary Accounting Research* 11(1): 223-250.
- Brown, L., and J. Gardner. 1985. "Applying citation analysis to evaluate the research contributions of accounting faculty and doctoral programs." *The Accounting Review* 60 (2): 262-277.
- Bujaki, M., and B. Mcconomy. 2017. "Productivity in top-10 academic accounting journals by researchers at Canadian universities at the start of the 21st century." *Accounting Perspectives* 16(4): 269-313.
- Chan, K. C., G. Seow, and K. Tam. 2009. "Ranking accounting journals using dissertation citation analysis: A research note," *Accounting, Organizations and Society* 34 (6): 875-885.

- Clarivate Analytics (2019). *Incites Journal Citation Reports*, April 21.
- Cready, W. 2019. "Complacency at the gates." *Significance* 16 (4): 18-19.
- Cready, W., J. He, W. Liu, C. Shao, D. Wang, & Y. Zhang. 2019. "Is there a confidence interval for that? A critical examination of null outcome reporting in accounting research." Working paper (July).
- Dyckman, T., & S. Zeff (1984). "Two decades of the Journal of Accounting Research." *Journal of Accounting Research* 22: 225-297.
- Ericson, B., & G. Brandt (2011) "High court rejects 'Statistical Significance' as materiality test for pharma securities fraud." <https://www.pillsburylaw.com/en/news-and-insights/high-court-rejects-statistical-significance-as-materiality-test.html>.
- Fidler, F., M. Burgman, G. Cumming, R. Buttrose, and N. Thomason. 2006. "Impact of criticism of null-hypothesis significance testing on statistical reporting practices in conservation biology." *Conservation Biology* 20 (5): 1539-1544.
- Financial Accounting Standards Board (2010). *Statement of Financial Accounting Concepts No. 8: Conceptual Framework for Financial Reporting Chapters 1 & 3*. Financial Accounting Foundation, Norwalk, CN
https://www.fasb.org/jsp/FASB/Document_C/DocumentPage?cid=1176157498129
- Hasselback, J., A. Reinstein, and M. Abdolmohammadi. 2012. "Benchmarking the research productivity of accounting doctorates." *Issues in Accounting Education* 27 (4): 943-78.
- Hoekstra, R., S. Finch, H. Kiers, and A. Johnson. 2006. "Probability as certainty: Dichotomous thinking and the misuse of *p* values." *Psychonomic Bulletin & Review* 13(6): 1033-1037.
- Howard, T.P., and L.A. Nikolai, 1983. "Attitude measurement and perceptions of accounting faculty publication outlets." *The Accounting Review* 61 (4): 735-743.
- Hull, R., & G. Wright (1990). "Faculty perceptions of journal quality: an update." *Accounting Horizons* 4(1): 77-98.
- Matherly, M., & R. Shortridge. 2009. "A pragmatic model to estimate journal quality in accounting." *Journal of Accounting Education* 27: 14-29.
- Matrixx Initiatives, Inc. vs. Siracusano. 563 U.S. 27 (2011).
- McRae, T. W. (1974). "A citational analysis of the accounting information network," *Journal of Accounting Research* 12(1): 80-92.

Schatz, P., K. Jay, J. McComb, and J. McLaughlin. 2005. "Misuse of statistical tests in *Archives of Clinical Neuropsychology* publications" *Archives of Clinical Neuropsychology* 20: 1053-1059.

SCImago, (2019). SJR—SCImago Journal & Country Rank. Retrieved 6/30/2019 from <https://www.scimagojr.com/journalrank.php?category=1402>

Tahai, A., & T. J. Rigsby, (1998). "Information processing using citations to investigate journal influence in accounting." *Information Processing and Management* 34(2/3), 341– 359.

Wasserstein, R.L., & N.A. Lazar (2016). "The ASA's Statement on *P*-values: Context, Process, and Purpose." *The American Statistician*, 70: 129-133.

Wasserstein, R.L., A.L. Schrim, & N.A. Lazar (2019). "Moving to a World Beyond " $p < .05$." *The American Statistician* 73.sup 1: 1-19.

Table 1

SCImago Journal Rank and JCR Journal Impact Factor Values for Publications From 2013-2017

Panel A: SCImago Journal Rank Values by Year

Journal	2018	2017	2016	2015	Rank Range
CAR	2.90	2.60	3.27	2.81	4-5
JAE	6.61	6.88	8.04	7.29	1-2
JAR	10.15	6.96	6.05	5.46	1-2
RAST	3.38	2.78	3.02	2.20	4-5
TAR	5.24	3.95	3.92	4.75	3

Panel B: JCR 2 Year Journal Impact Factor Values

Journal	2018	2017	2016	2015	Rank Range
CAR	2.26	2.07	2.27	1.78	4
JAE	3.75	3.28	3.84	3.54	1-3
JAR	4.89	4.54	3.00	2.24	1-2
RAST	2.11	1.59	1.76	1.51	5
TAR	4.56	2.25	2.30	1.95	2-3

Table 2
Cross-Journal Article Distribution by Topic
Articles Published Between 2013 and 2017.

This table reports article counts by topic area and journal for all articles published in journals (Panel A) and only those articles employing p-value based inference (Panel B). Topic area percentages by journal are reported in parentheses below each count.

Panel A: Distribution of All Articles						
Journal	Financial	Managerial	Audit	Tax	Other	Total
	Number (%)	Number (%)	Number (%)	Number (%)	Number (%)	Number
CAR	144 (52.9%)	47 (17.3%)	58 (21.3%)	18 (6.6%)	5 (1.8%)	272
JAE	126 (72.0%)	19 (10.9%)	15 (8.6%)	15 (8.6%)	0 (0.0%)	175
JAR	106 (68.4%)	19 (12.3%)	16 (10.3%)	10 (6.5%)	4 (2.6%)	155
RAST	163 (79.1%)	12 (5.8%)	17 (8.3%)	14 (6.8%)	0 (0.0%)	206
TAR	187 (51.9%)	61 (16.9%)	75 (20.8%)	33 (9.2%)	4 (1.1%)	360
Total	726 (62.2%)	158 (13.5%)	181 (15.5%)	90 (7.7%)	13 (1.1%)	1,168

Panel B: Distribution of Articles Employing Hypothesis Testing (p-value inference)						
Journal	Financial	Managerial	Audit	Tax	Other	Total
	Number (%)	Number (%)	Number (%)	Number (%)	Number (%)	Number
CAR	130 (56.0%)	37 (15.9%)	48 (20.7%)	16 (6.9%)	1 (0.4%)	232
JAE	115 (71.9%)	16 (10.0%)	14 (8.8%)	15 (9.4%)	0 (0.0%)	160
JAR	89 (67.9%)	15 (11.5%)	15 (11.5%)	10 (7.6%)	2 (1.5%)	131
RAST	149 (81.9%)	4 (2.2%)	16 (8.8%)	13 (7.1%)	0 (0.0%)	182
TAR	173 (53.2%)	46 (14.2%)	72 (22.2%)	31 (9.5%)	3 (0.9%)	325
Total	656 (63.7%)	118 (11.5%)	165 (16.0%)	85 (8.3%)	6 (0.6%)	1,030

Table 3
Cross-Tabulations by Journal, Research Area, and Method
Articles Published Between 2013 and 2017

This table presents cross-tabulations by research method and research area (panel A), and by research method and journal (Panel B) for articles employing p-value based inference. Research method percentages by research area (panel A) and journal (panel B) are reported in parentheses below each count value.

Panel A: Method Distribution by Research Area						
Method	Financial	Managerial	Audit	Tax	Other	Total
	Number (%)	Number (%)	Number (%)	Number (%)	Number (%)	Number
Archival	618 (94.2%)	82 (69.5%)	130 (79.3%)	84 (98.8%)	5 (83.3%)	918
Experimental	38 (5.8%)	36 (30.5%)	35 (20.7%)	1 (1.2%)	1 (16.7%)	110
Total	656	118	165	85	6	1,030

Panel B: Method Distribution by Journal						
Method	CAR	JAЕ	JAR	RAST	TAR	Total
	Number (%)	Number (%)	Number (%)	Number (%)	Number (%)	Number
Archival	198 (85.3%)	158 (98.75%)	112 (86.0%)	181 (99.5%)	270 (83.1%)	918
Experimental	34 (14.7%)	2 (1.25%)	19 (14.0%)	1 (0.5%)	55 (16.9%)	110
Total	232	160	131	182	325	1,030

Table 4

Null Outcome Misrepresentation Rates by Journal

This table reports percentages of articles with misleading abstract interpretations of null outcome evidence by journal. Total number of articles is the number of articles in a journal employing p-value based inference. The non-neutral/total values are percentages of total articles that employ non-neutral null outcome interpretations (broadly construed) in their abstracts. PC/total values are percentages of total articles that employ highly conclusive non-neutral null outcome interpretations (PC interpretations) in their abstracts. PC/Null-outcome-articles is the percentage of a journal's null outcome reporting articles that provide PC interpretations.

Journal	Total Number of Articles	Null Outcome Misrepresentation Rates		
		Non-Neutral/Total	PC/Total	PC/Null-Outcome-Articles
CAR	232	28.4%	17.7%	61.1%
JAE	160	19.4%	11.3%	58.1%
JAR	131	32.8%	24.4%	74.4%
RAST	182	24.2%	15.9%	65.9%
TAR	325	25.5%	16.9%	66.3%
ALL	1030	25.7%	17.0%	65.3%

Table 5

Null Outcome Misrepresentation Rates by Research Area and Method

This table reports percentages of articles with misleading abstract interpretations of null outcome evidence by research area (panel A) and method (panel B). Total number of articles is the number of articles in an area or method classification employing p-value based inference. The non-neutral/total values are percentages of total articles that employ non-neutral null outcome interpretations (broadly construed) in their abstracts. PC/total values are percentages of total articles that employ highly conclusive non-neutral null outcome interpretations (PC interpretations) in their abstracts. PC/Null-outcome-articles is the percentage of null outcome reporting articles in an area or method classification that provide PC interpretations.

Panel A: Null Outcome Misrepresentation by Research Area

Sub-Discipline	Total Number of Articles	Null Outcome Misrepresentation Rates		
		Non-Neutral/Total	PC/Total	PC/Null-Outcome-Articles
Financial	656	23.2%	14.5%	62.5%
Managerial	119	26.1%	18.5%	68.8%
Audit	164	35.4%	24.4%	68.9%
Tax	85	28.3%	20.0%	70.8%
TOTAL	1024	25.9%	17.0%	65.4%

Panel B: Null Outcome Misrepresentation by Research Method

Method	Total Number of Articles	Null Outcome Misrepresentation Rates		
		Non-Neutral/Total	PC/Total	PC/Null-Outcome-Articles
Archival	919	25.9%	16.3%	62.8%
Experimental	111	26.1%	22.5%	86.2%
TOTAL	1030	25.9%	17.0%	65.3%

Table 6

Null Outcome Misrepresentation Rates for Financial Archival Articles

This table reports percentages of financial archival articles with misleading abstract interpretations of null outcome evidence by journal. Total number articles is the number of financial archival articles in a journal employing p-value based inference. The non-neutral/total values are percentages of total articles that employ non-neutral null outcome interpretations (broadly construed) in their abstracts. PC/total values are percentages of total articles that employ highly conclusive non-neutral null outcome interpretations (PC interpretations) in their abstracts. PC/Null-outcome-articles is the percentage of a journal's financial archival null outcome reporting articles that provide PC interpretations.

Journal	Total Number of Articles	Null Outcome Misrepresentation Rates		
		Non-Neutral/Total	PC/Total	PC/Null-Outcome-Articles
CAR	118	29.7%	16.9%	55.6%
JAE	113	17.7%	10.6%	60.0%
JAR	81	30.9%	22.2%	72.0%
RAST	148	23.0%	14.2%	61.8%
TAR	158	20.3%	12.0%	59.4%
ALL	618	25.2%	14.6%	57.3%

Table 7

Article Null Outcome Reporting Quality Assessments Over Time

This table reports percentages of articles with misleading abstract interpretations of null outcome over time. Total number of articles is the number of articles in a journal employing p-value based inference. The non-neutral/total values are percentages of total articles that employ non-neutral null outcome interpretations (broadly construed) in their abstracts. PC/total values are percentages of total articles that employ highly conclusive non-neutral null outcome interpretations (PC interpretations) in their abstracts. PC/Null-outcome-articles is the percentage of null outcome reporting articles in a given year that provide PC interpretations.

Year	Total Number of Articles	Null Outcome Misrepresentation Rates		
		Non-Neutral/Total	PC/Total	PC/Null-Outcome-Articles
2013	205	26.8%	19.5%	72.8%
2014	189	23.8%	15.3%	64.4%
2015	209	27.3%	18.2%	66.7%
2016	208	24.5%	15.4%	62.7%
2017	219	26.0%	16.4%	63.2%
ALL	1030	25.7%	17.0%	66.0%